Diabetes Diagnoser

As the title suggests, this project is an end-to-end example of solving a real-world problem using Data Science. We’ll be using Machine Learning to predict whether a person has diabetes or not, based on information about the patient such as blood pressure, body mass index (BMI), age, etc. The project walks through the various stages of the data science workflow. In particular, the project has the following sections:

* Project Overview
* Project Description
* Project Preparation
* Project Dependencies
* Project Responsiveness

Overview

The data was collected and made available by “National Institute of Diabetes and Digestive and Kidney Diseases” as part of the Pima Indians Diabetes Database. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here belong to the Pima Indian heritage (subgroup of Native Americans), and are females of ages 21 and above.

Description

The following features have been provided to help us predict whether a person is diabetic or not:

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration over 2 hours in an oral glucose tolerance test

Blood Pressure: Diastolic blood pressure (mm Hg)

Skin Thickness: Triceps skin fold thickness (mm)

Insulin:2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)2)

Diabetes Pedigree Function: Diabetes pedigree function (a function which scores likelihood of diabetes based on family history)

Age: Age (years)

Class variable (0 if non-diabetic, 1 if diabetic)

Preparation

The project is prepared by training a modal using logistic regression.

The following modules are used for training the modal:

* **Numpy**

Converting and collecting the raw data into arrays like structure for model is done by numpy.

* **Pandas**

The module pandas is used for loading a csv file in which the raw data was uploaded by the institute. Also relocating the data into dataframes and exploration of the dataflow is done by pandas

* **Pickle**

Pickle does a simple job of converting a model into a dump file which is used to connect to the app file for user and frontend.

* **Sklearn.linear\_model**

Logistic Regression is a part of Sklearn.linear\_model

* **Sklearn.impute**

An imputer is used for imputing the missing data in diabetes dataset.

* **Sklearn.pipeline**

To increase the throughput of model, pipeline is used.

Dependencies

A series of interrelated tasks are done in the project, python-Flask is used as a backend and frontend manager with the help of Jinja2 templating and SQLAlchemy.  
HTML, CSS and JavaScript is used as frontend languages for developing GUI for project

User Responsiveness

The primary focus of the project is user responsiveness .Without users the product is good for nothing.

Developing a machine learning model is hard but creating a user responsive model is knotty and complex. To handle such complex problems we have used flask for both frontend and backend.